Mark Villar
Information and Communications Technology
University of Sydney

Aim

This study explores the socioeconomic circumstances and public health outcomes of Indigenous Australians, with particular emphasis on the quantitative analysis of large data sets. Our overarching aim is to provide better health and welfare services to Indigenous Australians, through the implementation of effective public policy. The Indigenous Health Check (MBS 715) data tool, provided by the Australian Institute of Health and Welfare (AIHW), is our primary source of raw data. Using the R and ArcGIS platforms, we hope to transform the existing SAS-based application to improve not just its data visualization features and mobile capabilities, but more importantly, the reliability of information obtained by government decision-makers in trying to 'close the gap' between Indigenous and non-Indigenous Australians.

Data

The Health and Welfare of Australia's Aboriginal and Torres Strait Islander People: An Overview (2011)

A comprehensive 127-page report outlining the following:

- Demographics of Aboriginal and Torres Strait Islander people
- Determinants of health and welfare: socioeconomic factors, housing, community capacity, behavioural factors, social and emotional wellbeing
- Health and functioning: community functioning, disability, health conditions
- Mortality and life expectancy
- Health across the life stages: mothers and babies, children, young people, older people
- Health care and other support services
- Health and welfare expenditure

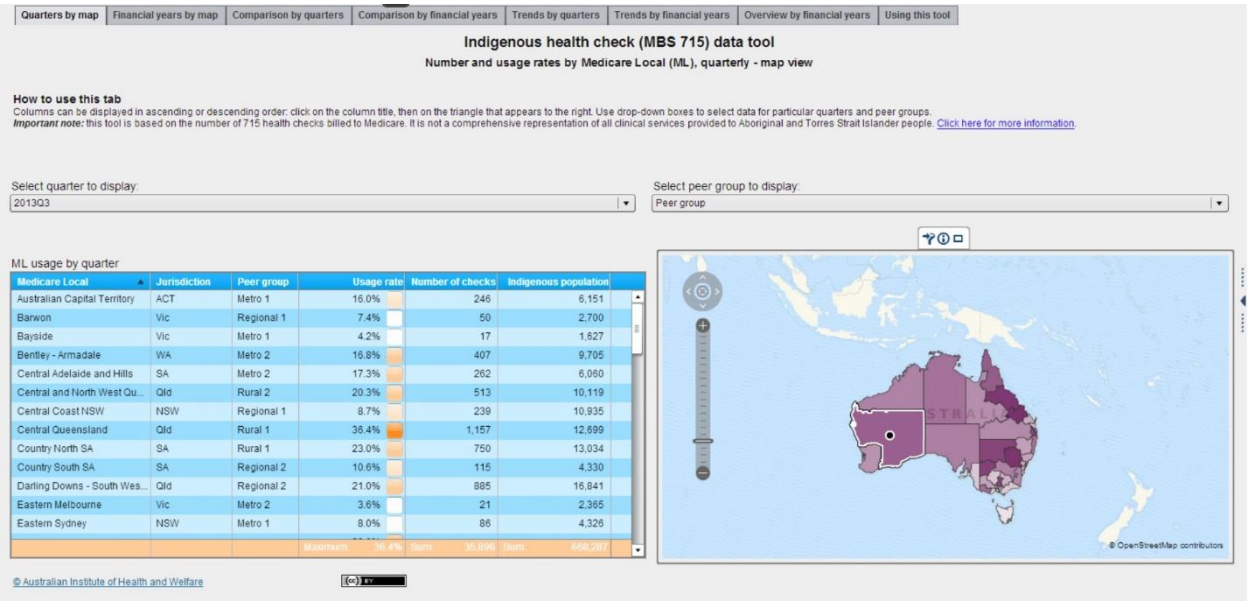
Source data as an Excel download (200 KB XLS)

Medicare Benefits Schedule (MBS) 715 is an annual health check for Aboriginal and Torres Strait Islander people of all ages.

- Table 1: MBS item 715 by Medicare Local by quarter
- Table 2: MBS item 715 by Medicare Local by financial year

Indigenous Health Check (MBS 715) data tool: *SAS interactive data portal*

- Tab 1: Quarters by map
- Tab 2: Financial years by map
- Tab 3: Comparison by quarters
- Tab 4: Comparison by financial years
- Tab 5: Trends by quarters
- Tab 6: Trends by financial years
- Tab 7: Overview by financial years
- Tab 8: Using this tool



Specifications

- 61 Medicare Locals (MLs) – replaced *Divisions of General Practice* in 2011; MLs vary in terms of size, remoteness and population characteristics which makes comparisons between them difficult.
- 7 Peer groups – based on socioeconomic status and remoteness, including average distance to the closest large capital city and major hospital, making this type of stratification better for comparisons. The National Health Performance Authority allocated each ML to one of seven peer groups: Metro 1, Metro 2, Metro 3, Regional 1, Regional 2, Rural 1, Rural 2
- 8 Jurisdictions – ACT, NSW, NT, QLD, SA, TAS, VIC, WA
- ML Code – unique three-digit code for each of the 61 MLs.

- Usage rate (quarterly) = number of 715 health checks in the quarter divided by a quarter of the Indigenous population in the ML.
- Usage rate (financial year) = number of 715 health checks in the financial year divided by the Indigenous population in the ML.
- Number of checks = number of Indigenous health checks billed to Medicare as MBS item 715
- Indigenous population – data sourced from Population Health Information Development Unit at the University of Adelaide

Exploratory Data Analysis

Table 1: MBS item 715 by ML by quarter

- 488 observations of 7 variables
- Missing values (n=6) – Number of checks and Usage rates are not available for 6 observations where Indigenous populations are quite small (approx. 1300-2500 people). Three of the six missing values belong to Metro 1 groups in either NSW or VIC (ML107, ML108, ML206). Meanwhile, there are two missing quarterly observations for the same Metro 2 group in VIC (ML203) and another in a Regional 1 group, also in VIC (ML209).

Table 2: MBS item 715 by ML by financial year

- 122 observations of 7 variables
- No missing values in this table.

Variables

- *jurisd* – factor with 8 levels (nominal)
- *mlcode* – factor with 61 levels (nominal)
- *peergrp* – factor with 7 levels (nominal/ordinal)
- *quarter* – factor with 9 levels (ordinal)
- *year* – factor with 2 levels (ordinal)
- *nchecks* – number of checks (count)
- *usage* – usage rate (proportion)
- *popn* – population (count)

SAS interactive data tool – Indigenous Health Check (MBS 715)

The following Indigenous populations were obtained using the pull-down menus under Tabs 1 & 3 of the web-based data tool provided by the AIHW.

Peer group	Population
Metro 1	24,798
Metro 2	94,295
Metro 3	63,020
Regional 1	81,031
Regional 2	172,791
Rural 1	65,110
Rural 2	167,242
Total	**668,287**

Jurisdiction	Population
NSW	204,533
VIC	50,999
QLD	188,503
SA	37,397
WA	87,693
TAS	24,165
ACT	6,151
NT	68,846
Total	**668,287**

Jurisdiction	Metro 1	Metro 2	Metro 3	Regional 1	Regional 2	Rural 1	Rural 2	Total
NSW	13,586	3,652	29,980	62,901	86,382	8,033	0	204,534
VIC	5,061	4,325	11,034	5,207	21,663	3,709	0	50,999
QLD	0	46,293	12,920	7,051	25,267	40,334	56,638	188,503
SA	0	10,948	9,086	0	4,330	13,034	0	37,398
WA	0	29,077	0	5,872	10,986	0	41,758	87,693
TAS	0	0	0	0	24,165	0	0	24,165
ACT	6,151	0	0	0	0	0	0	6,151
NT	0	0	0	0	0	0	68,846	68,846
Total	24,798	94,295	63,020	81,031	172,793	65,110	167,242	**668,289**

We also collected data on the *maximum usage rates* and *total number of checks* for both quarterly and yearly intervals from Tabs 1 & 2. Further breakdowns that compare the data by *peer group* and *jurisdiction* were also obtained from Tabs 3 & 4. The tables below summarise the data we extracted from the interactive data tool.

Tab 1: Quarters by map: number of checks and usage rates by ML (map view)

- White geospatial map
- 9 quarters – 2011Q3 to 2013Q3

Quarter	Max usage rate (%)	Total number of checks
2011Q3	38.6	21,990
2011Q4	32.0	21,721
2012Q1	35.0	25,245
2012Q2	36.3	27,564
2012Q3	36.2	28,920
2012Q4	45.5	28,085
2013Q1	32.9	30,309
2013Q2	39.0	34,813
2013Q3	36.4	35,896

Tab 2: Financial years by map: number of checks and usage rates by ML (map view)

- Greyscale geospatial map
- 2 financial years – 2011-2012 and 2012-2013

Financial year	Max usage rate (%)	Total number of checks
2011-12	32.3	96,520
2012-13	34.1	122,127

Tab 3: Comparison by quarters: number of checks and usage rates by ML (bar chart view)

- User selects Jurisdiction, Peer group and Quarter to display Usage rate information as a *static bar chart*

Quarter	Metro 1		Metro 2		Metro 3		Regional 1		Regional 2		Rural 1		Rural 2	
	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks
2011Q3	8.3	321	18.0	2,523	18.8	1,458	9.5	1,297	24.6	5,282	38.6	3,608	21.4	7,501
2011Q4	11.6	467	15.8	2,403	14.8	1,229	13.1	1,421	32.0	5,739	31.0	3,649	19.1	6,813
2012Q1	10.2	461	22.6	3,144	19.0	1,448	14.2	1,629	35.0	6,202	32.4	4,383	20.2	7,978
2012Q2	13.0	483	21.6	3,205	16.0	1,411	14.4	2,115	31.3	6,719	36.3	4,712	25.9	8,919
2012Q3	14.2	520	22.4	3,766	20.2	1,717	16.6	2,217	30.5	6,873	36.2	4,963	27.2	8,864
2012Q4	18.5	563	22.7	3,895	14.1	1,431	16.7	2,064	45.5	6,588	32.8	4,276	24.8	9,268
2013Q1	14.6	583	29.5	4,724	16.3	1,489	15.9	2,081	29.6	7,124	32.9	4,382	28.6	9,926
2013Q2	17.6	594	35.1	5,166	19.3	1,878	17.1	2,467	35.0	8,453	39.0	5,225	34.4	11,030
2013Q3	16.0	634	34.2	5,416	31.7	2,417	17.8	2,828	35.7	8,691	36.4	5,314	29.0	10,596
Total		**4,626**		**34,242**		**14,478**		**18,119**		**61,671**		**40,512**		**80,895**
Popn		24,798		94,295		63,020		81,031		172,791		65,110		167,242

Quarter	ACT		NSW		NT		QLD		SA		TAS		VIC		WA	
	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks	Max usage rate	Number of checks
2011Q3	3.4	53	38.6	6,289	19.0	3,268	23.5	8,431	9.9	663	2.4	144	18.7	922	16.1	2,220
2011Q4	11.6	179	31.0	6,391	19.1	3,290	27.1	7,843	9.1	598	2.6	160	32.0	1,153	13.5	2,107
2012Q1	7.3	113	28.2	6,966	20.0	3,442	32.4	9,557	16.1	863	3.6	220	35.0	1,209	19.5	2,875
2012Q2	11.2	172	31.6	7,769	21.5	3,701	36.3	10,824	15.9	874	4.2	252	31.3	1,237	18.4	2,735
2012Q3	14.2	218	31.2	8,129	22.3	3,840	36.2	11,275	17.4	965	4.1	249	27.1	1,189	21.3	3,055
2012Q4	18.5	284	32.8	7,601	23.3	4,010	30.9	9,833	16.0	931	3.8	227	45.5	1,476	24.8	3,723
2013Q1	11.8	181	26.2	7,823	25.6	4,407	32.9	11,363	12.7	842	5.3	322	29.6	1,433	28.6	3,938
2013Q2	12.0	185	39.0	9,403	23.8	4,104	36.1	13,165	17.6	1,142	5.9	358	34.1	1,655	35.1	4,801
2013Q3	16.0	246	35.7	9,807	25.3	4,351	36.4	13,357	23.0	1,372	5.6	340	33.4	1,796	30.5	4,627
Total		**1,631**		**70,178**		**34,413**		**95,648**		**8,250**		**2,272**		**12,070**		**30,081**
Popn		6,151		204,533		68,846		188,503		37,397		24,165		50,999		87,693

Tab 4: Comparison by financial years: number of checks and usage rates by ML (bar chart view)

- User selects Jurisdiction, Peer group and Year to display Usage rate information as a *static bar chart*

Year	Metro 1		Metro 2		Metro 3		Regional 1		Regional 2		Rural 1		Rural 2	
	Max usage rate	**Number of checks**	**Max usage rate**	**Number of checks**	**Max usage rate**	**Number of checks**	**Max usage rate**	**Number of checks**	**Max usage rate**	**Number of checks**	**Max usage rate**	**Number of checks**	**Max usage rate**	**Number of checks**
2011-12	9.9	1,732	17.8	11,275	17.1	5,546	12.7	6,462	28.0	23,942	32.3	16,352	21.0	31,211
2012-13	14.1	2,260	25.3	17,551	17.5	6,515	16.6	8,829	34.1	29,038	33.9	18,846	29.9	39,088
Total		**3,992**		**28,826**		**12,061**		**15,291**		**52,980**		**35,198**		**70,299**
Popn		24,798		94,295		63,020		81,031		172,791		65,110		167,242

Tab 5: Trends by quarters: usage rates by ML (trend line view)

- User selects ML, Jurisdiction, Peer group and Quarter(s).
- Tick boxes allow multiple selection by Quarters.
- **Note:** Quarterly trend lines can show considerable variations in usage. Annual reports give less variable information about usage. For this reason, we do not analyze this tab in further detail. Trends should also be interpreted with caution, particularly where they relate to small populations.

Tab 6: Trends by financial years: usage rates by ML (bar chart view)

- User selects ML, Jurisdiction, Peer group and Financial Year(s).

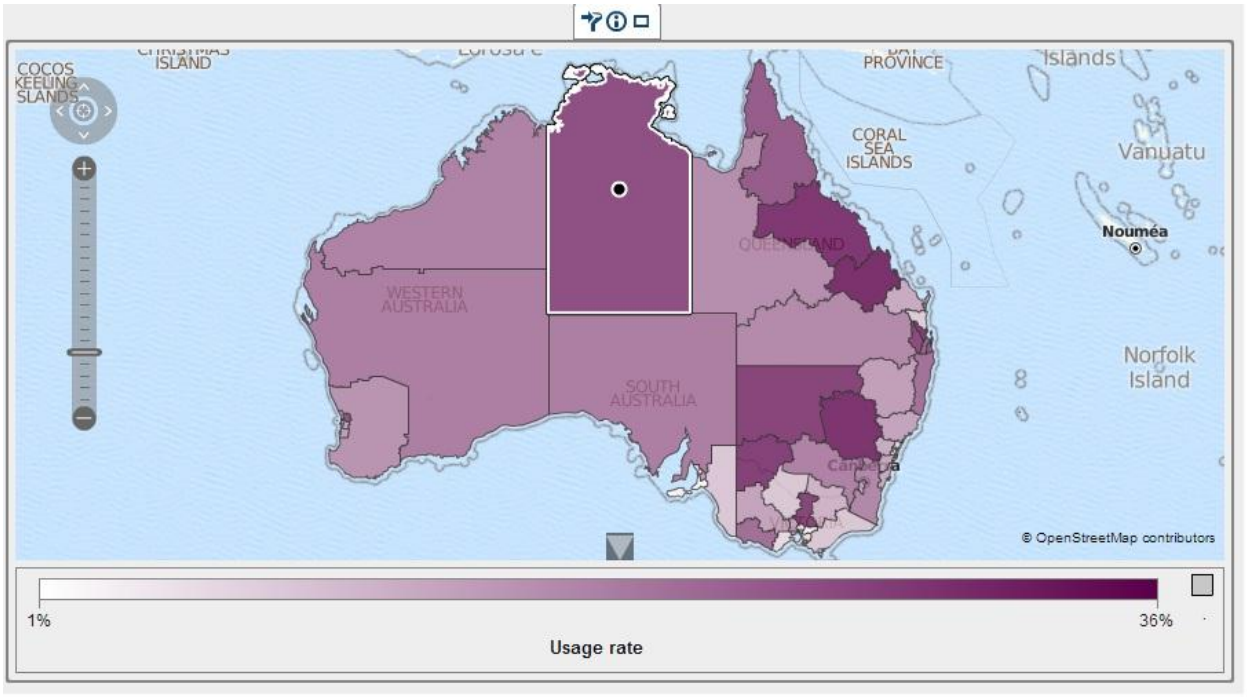
Tab 7: Overview by financial years: usage rates and population by ML (tree map view)

- User selects Jurisdiction, Peer group and Financial Year(s).

Tab 8: Using this tool: tips on using the tool and explanation of the terminology

Design

Interactive map below is totally underwhelming.

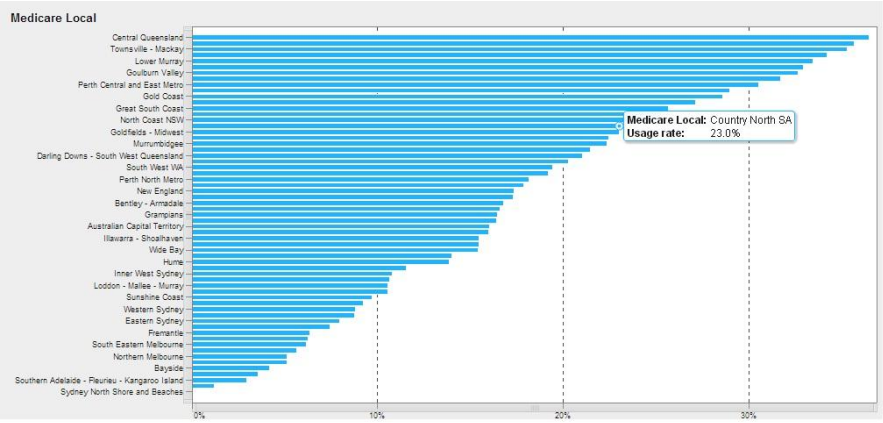


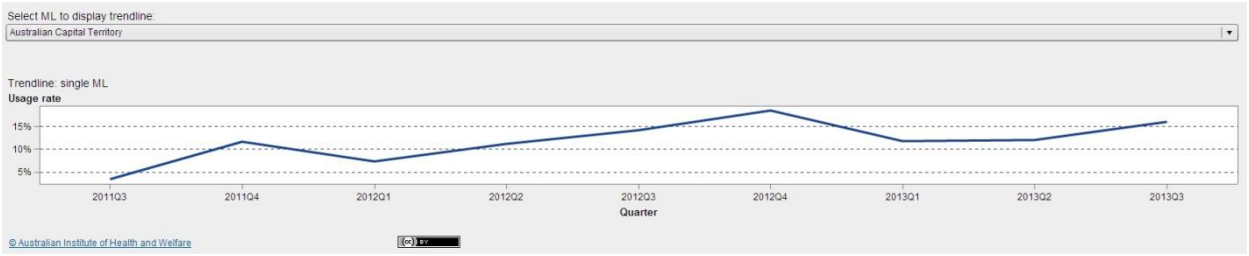
User interface with multiple tabs and pull-down menus – information is "hidden" and navigation is clunky.

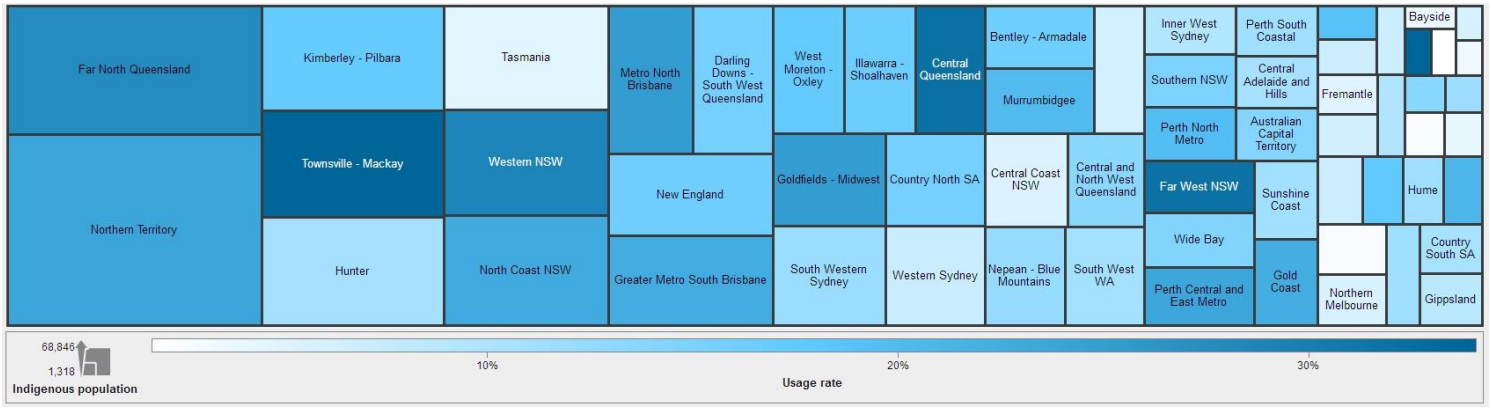
Colour scheme – poor text contrast makes Summary Statistics at the bottom of tables difficult to read clearly.

Medicare Local	Jurisdiction	Peer group	Usage rate	Number of checks	Indigenous population
Country North SA	SA	Rural 1	23.0%	750	13,034
Country South SA	SA	Regional 2	10.6%	115	4,330
Darling Downs - South Wes...	Qld	Regional 2	21.0%	885	16,841
Eastern Melbourne	Vic	Metro 2	3.6%	21	2,365
Eastern Sydney	NSW	Metro 1	8.0%	86	4,326
Far North Queensland	Qld	Rural 2	29.0%	3,367	46,519
Far West NSW	NSW	Rural 1	32.9%	661	8,033
Frankston - Mornington Pen...	Vic	Regional 1	5.1%	32	2,507
Fremantle	WA	Metro 2	6.3%	53	3,354
Gippsland	Vic	Regional 2	10.5%	122	4,637
Gold Coast	Qld	Metro 2	28.6%	557	7,802
Goldfields - Midwest	WA	Rural 2	23.0%	852	14,810
Goulburn Valley	Vic	Regional 2	32.6%	313	3,837
Grampians	Vic	Regional 2	16.5%	117	2,843
Great South Coast	Vic	Regional 2	25.7%	110	1,714
Greater Metro South Brisba...	Qld	Metro 2	27.1%	1,414	20,862
Hume	Vic	Regional 2	13.8%	132	3,818
			Maximum 36.4%	Sum 35,896	Sum 658,287

Static monochromatic charts – bar charts, trendlines, tree map (boring!!)







Hypothesis Testing

[Insert hypotheses here]

Methods

Geographical data tools:

http://www.r-bloggers.com/starting-analysis-and-visualisation-of-spatial-data-with-r/

http://www.r-bloggers.com/3d-mapping-in-r/

http://www.esri.com/software/arcgis/arcgisonline

https://sydneyuni.maps.arcgis.com

Statistical techniques:

- Analysis of variance (ANOVA)
- Contingency tables (chi-squared tests)
- Estimation and prediction (regression)
- Data mining and machine learning
- Factor and cluster analysis

Software

RStudio

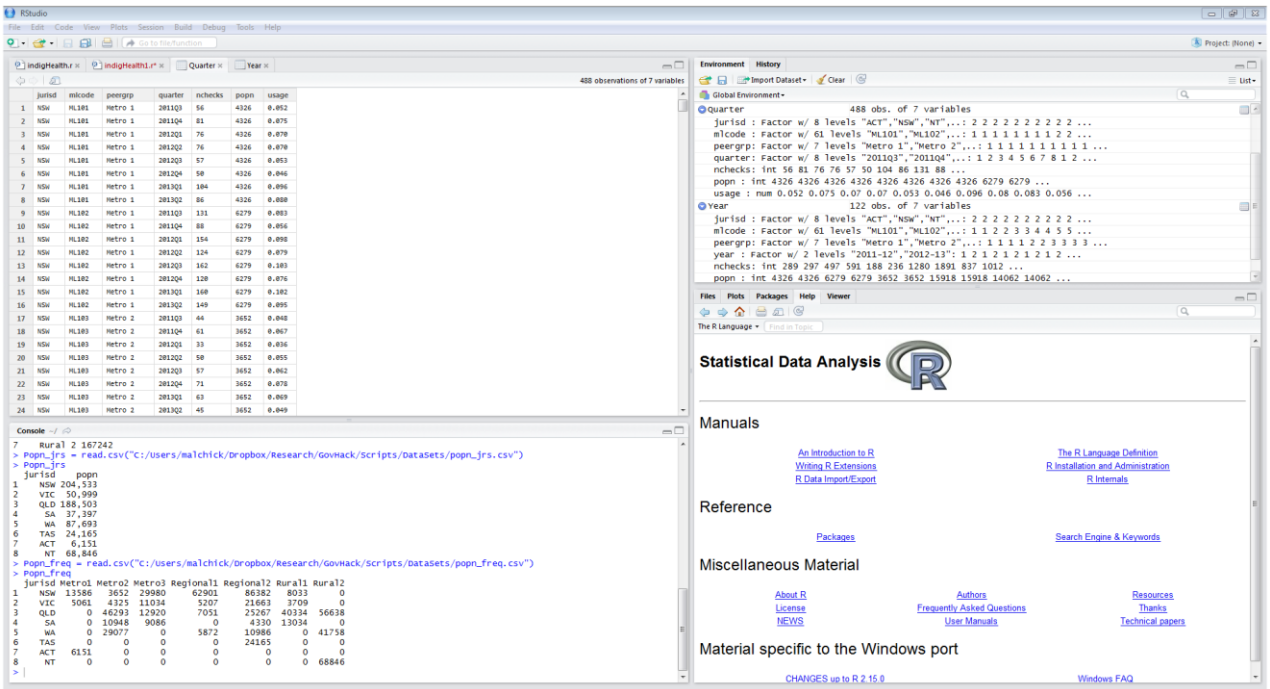
Statwing

ArcGIS

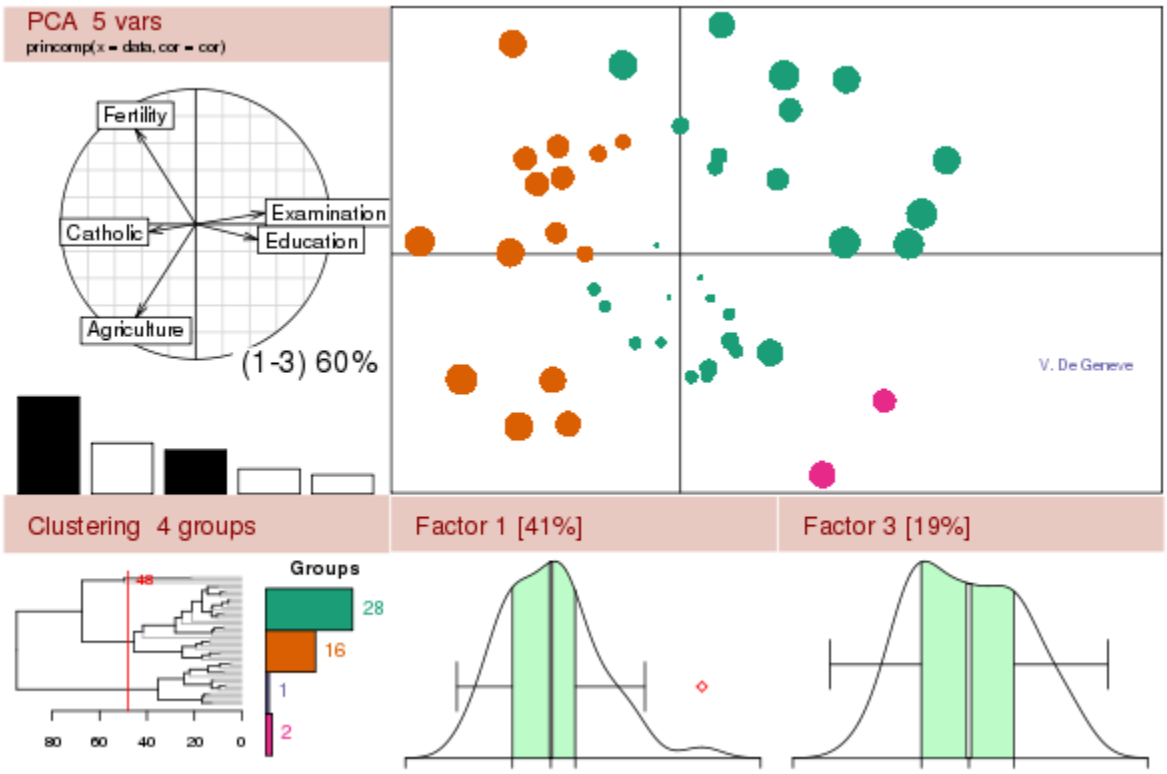
CartoDB

GitHub

ShareLaTeX



The R Project for Statistical Computing



Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.
- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News:

- **R 3.1.1 (Sock it to Me) prerelease versions** will appear starting June 30. Final release is scheduled for July 10, 2014.
- **R version 3.1.0** (Spring Dance) has been released on 2014-04-10.
- **R version 3.0.3** (Warm Puppy) has been released on 2014-03-06.
- **The R Journal Vol.5/2** is available.
- **useR! 2013**, took place at the University of Castilla-La Mancha, Albacete, Spain, July 10-12 2013.
- **R version 2.15.3** (Security Blanket) has been released on 2013-03-01.

Inspired by R and its community

The RStudio team contributes code to many R packages and projects. R users are doing some of the most innovative and important work in science, education, and industry. It's a daily inspiration and challenge to keep up with the community and all it is accomplishing.

ggplot2

An enhanced data visualization package for R. Create stunning multi-layered graphics with ease.

Project Site Link ›

knitr

Elegant, flexible and fast dynamic report generation that combines R with TeX, Markdown, or HTML.

Project Site Link ›

dplyr

dplyr is the next iteration of plyr, focussing on only data frames. dplyr is faster and has a more consistent API.

Project GitHub Link ›

R Markdown

R Markdown lets you insert R code into a markdown document. R then generates a final document that replaces the R code with its results.

Project Site Link ›

devtools

Developer tools for building R packages. Remove the pains and bottlenecks of package development.

Project Site Link ›

packrat

A dependency management tool for R to make your R projects more isolated, portable, and reproducible.

Project GitHub Link ›

ggvis

ggvis is the next iteration of the popular ggplot2 graphics package. ggvis creates dynamic, interactive data visualizations.

Project Site Link ›

plyr

An R package for efficiently manipulating data. Utilize the split, apply, combine method for fast, consistent results.

Project Site Link ›

reshape2

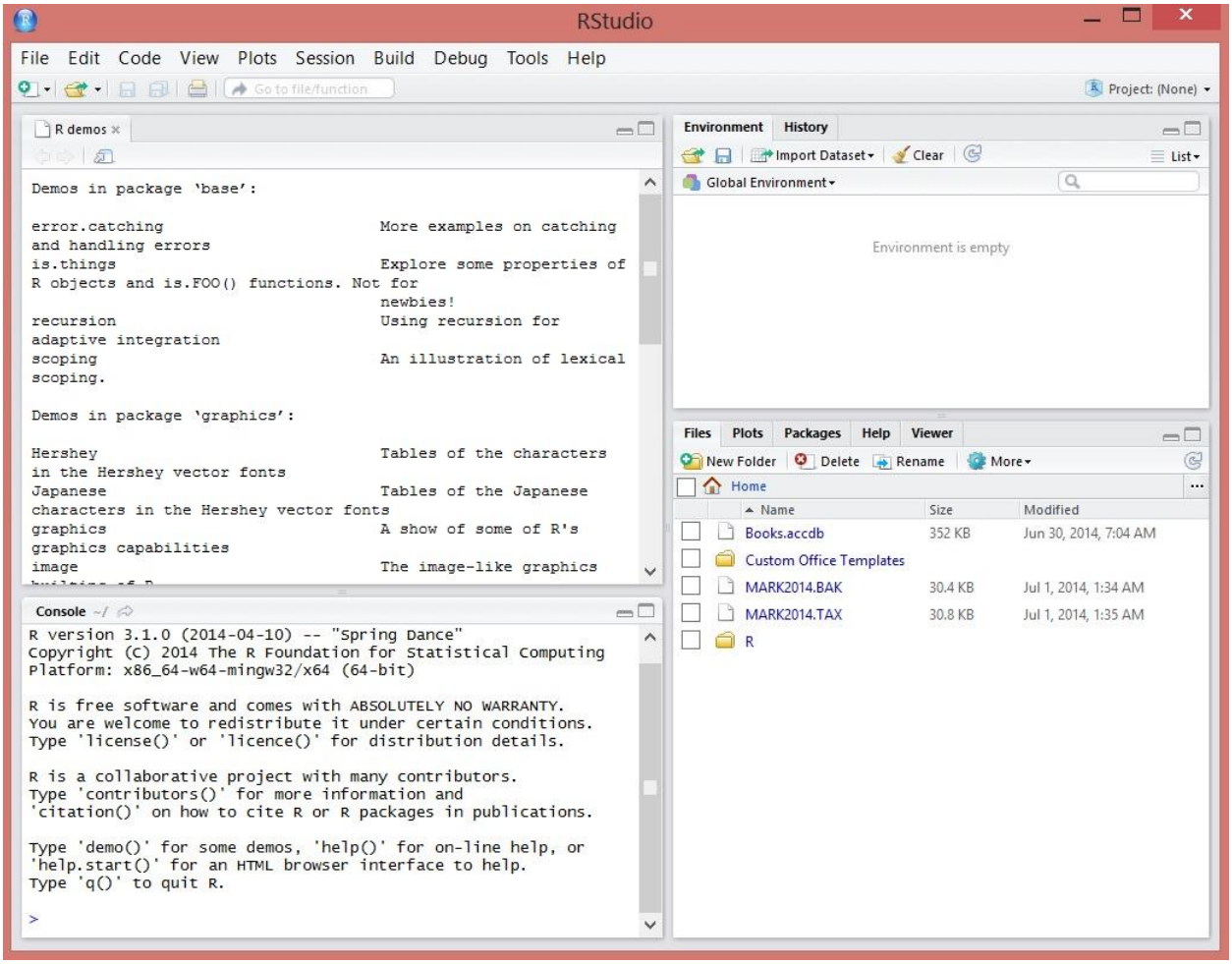
reshape2 makes it easy to transform data between wide and long formats. reshape2 is based around two key functions: melt and cast: melt takes wide-format data and melts it into long-format data and cast takes long-format data and casts it into wide-format data.

Project GitHub Link ›

shiny

Shiny makes it incredibly easy to build interactive web applications with R. Automatic "reactive" binding between inputs and outputs and extensive pre-built widgets make it possible to build beautiful, responsive, and powerful applications with minimal effort.

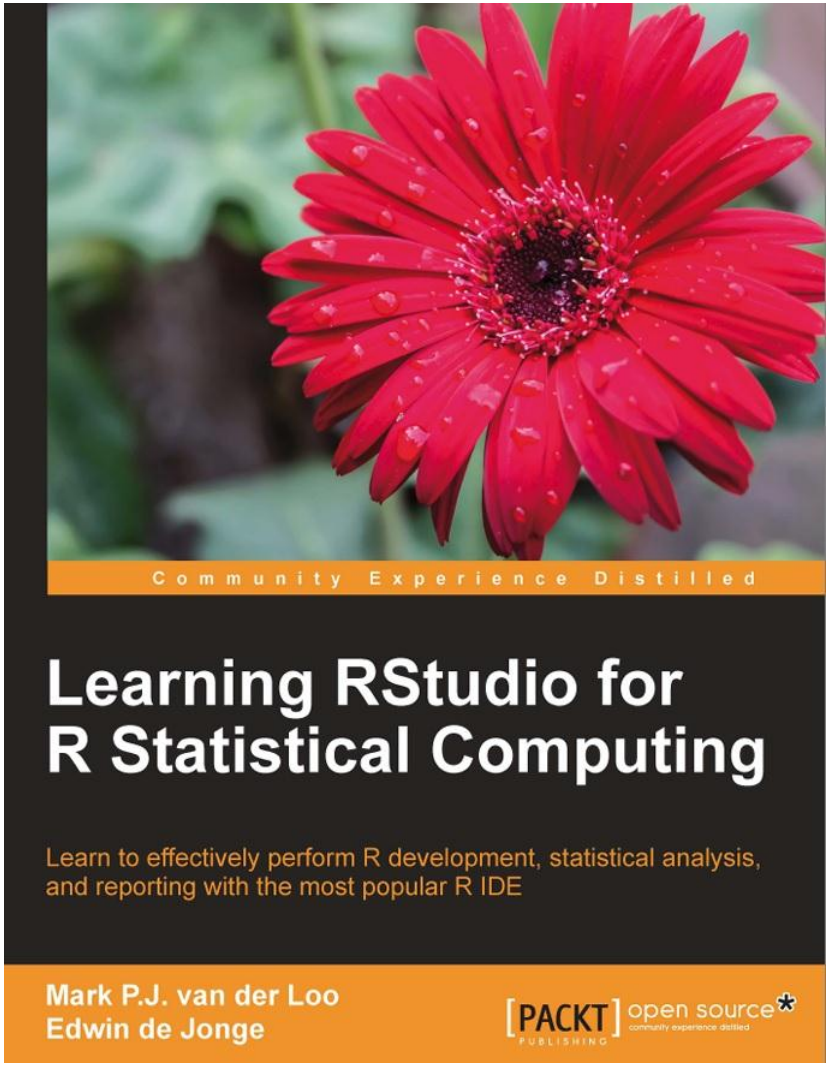
Project CRAN Link ›

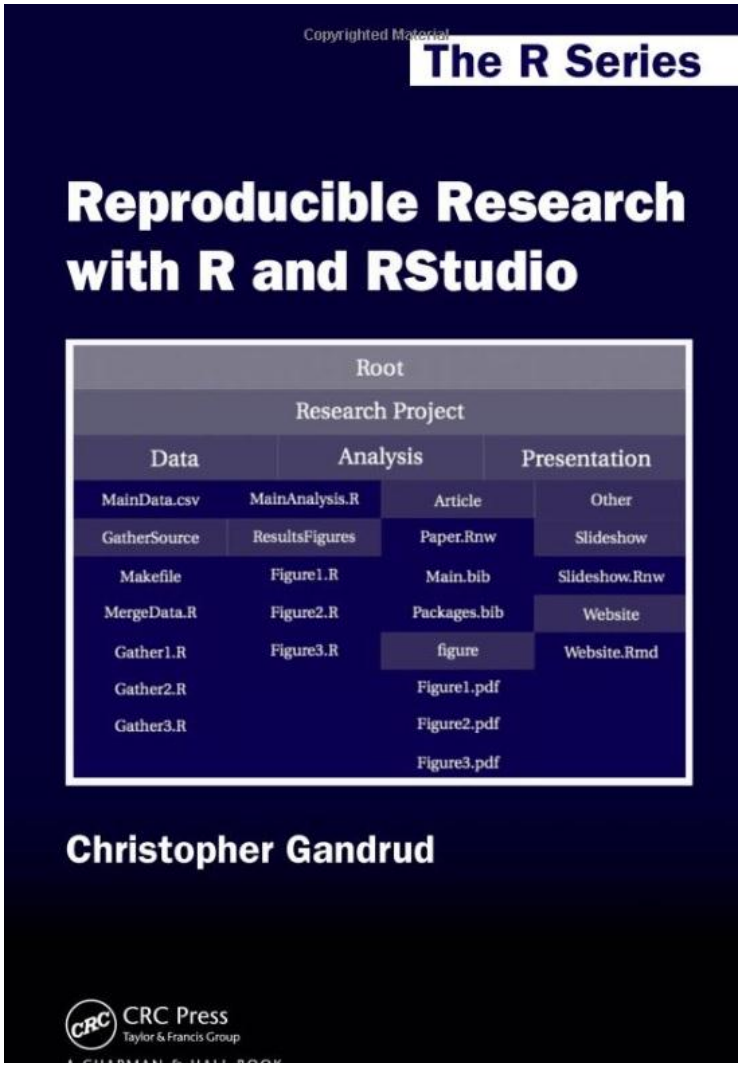


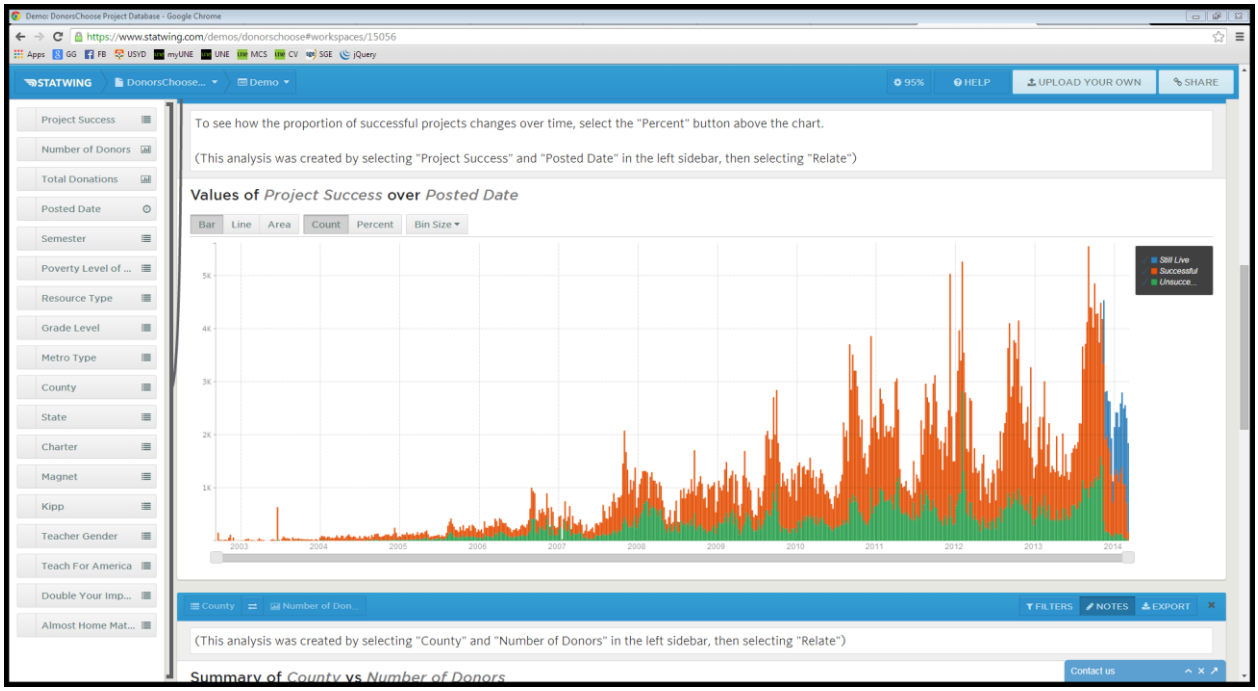
Learning RStudio for R Statistical Computing

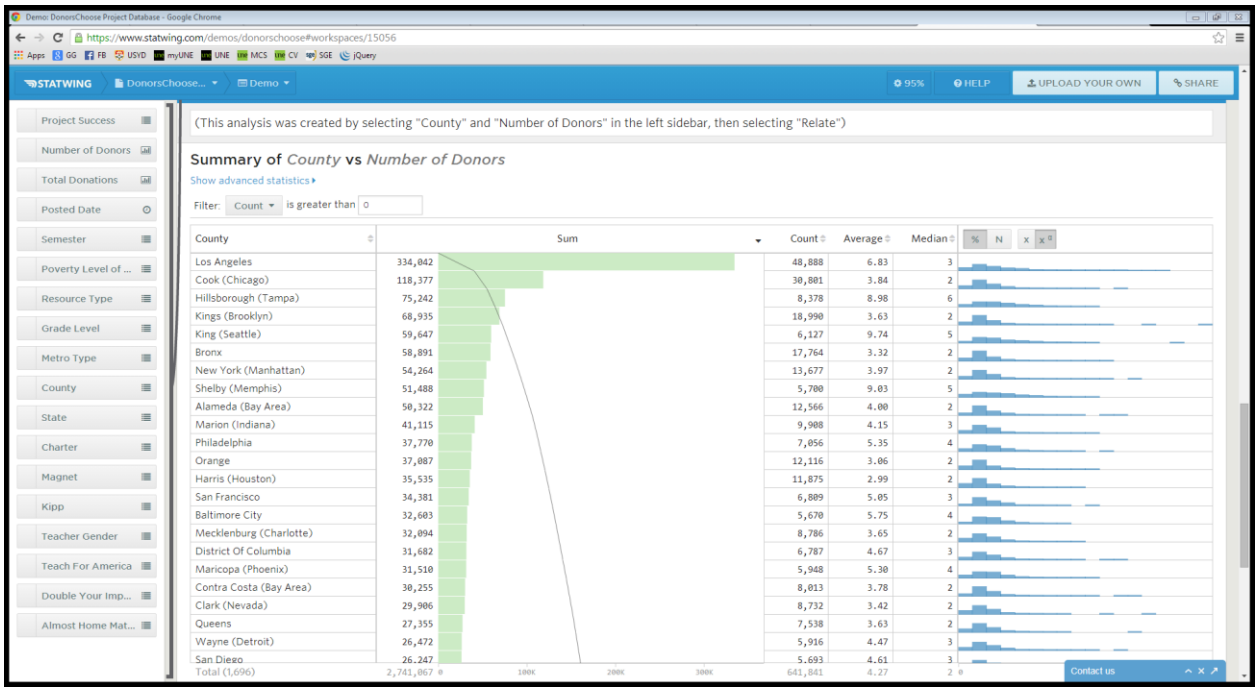
van der Loo, Mark P.J.; de Jonge, Edwin

A practical tutorial covering how to leverage RStudio functionality to effectively perform R Development, analysis, and reporting with RStudio. The book is aimed at R developers and analysts who wish to do R statistical development while taking advantage of RStudio functionality to ease their development efforts. Familiarity with R is assumed. Those who want to get started with R development using RStudio will also find the book useful. Even if you already use R but want to create reproducible statistical analysis projects or extend R with self-written packages, this book shows how to quickly achieve this using RStudio.


Community Experience Distilled
Learning RStudio for R Statistical Computing
Learn to effectively perform R development, statistical analysis, and reporting with the most popular R IDE
Mark P.J. van der Loo
Edwin de Jonge
[PACKT] open source
PUBLISHING


Copyrighted Material
The R Series
Reproducible Research with R and RStudio
Root
Research Project
Data
Analysis
Presentation
MainData.csv
GatherSource
Makefile
MergeData.R
Gather1.R
Gather2.R
Gather3.R
MainAnalysis.R
ResultsFigures
Figure1.R
Figure2.R
Figure3.R
Article
Paper.Rnw
Main.bib
Packages.bib
figure
Figure1.pdf
Figure2.pdf
Figure3.pdf
Other
Slideshow
Slideshow.Rnw
Website
Website.Rmd
Christopher Gandrud
CRC Press
Taylor & Francis Group


Demo: DonorsChoose Project Database - Google Chrome
https://www.statwing.com/demos/donorschoose#workspaces/15056
STATWING
DonorsChoose...
Demo
95%
HELP
UPLOAD YOUR OWN
SHARE
Project Success
Number of Donors
Total Donations
Posted Date
Semester
Poverty Level of ...
Resource Type
Grade Level
Metro Type
County
State
Charter
Magnet
Kipp
Teacher Gender
Teach For America
Double Your Imp...
Almost Home Mat...
To see how the proportion of successful projects changes over time, select the "Percent" button above the chart.
(This analysis was created by selecting "Project Success" and "Posted Date" in the left sidebar, then selecting "Relate")
Values of Project Success over Posted Date
Bar
Line
Area
Count
Percent
Bin Size
Still Live
Successful
Unsucce...
5K
4K
3K
2K
1K
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
County
Number of Don...
FILTERS
NOTES
EXPORT
(This analysis was created by selecting "County" and "Number of Donors" in the left sidebar, then selecting "Relate")
Summary of County vs Number of Donors
Contact us


Demo: DonorsChoose Project Database - Google Chrome
https://www.statwing.com/demos/donorschoose#workspaces/15056
STATWING
DonorsChoose...
Demo
95%
HELP
UPLOAD YOUR OWN
SHARE
Project Success
Number of Donors
Total Donations
Posted Date
Semester
Poverty Level of ...
Resource Type
Grade Level
Metro Type
County
State
Charter
Magnet
Kipp
Teacher Gender
Teach For America
Double Your Imp...
Almost Home Mat...
(This analysis was created by selecting "County" and "Number of Donors" in the left sidebar, then selecting "Relate")
Summary of County vs Number of Donors
Show advanced statistics
Filter: Count is greater than 0
County
Sum
Count
Average
Median
% N x x²
Los Angeles 334,042 48,888 6.83 3
Cook (Chicago) 118,377 30,801 3.84 2
Hillsborough (Tampa) 75,242 8,378 8.98 6
Kings (Brooklyn) 68,935 18,990 3.63 2
King (Seattle) 59,647 6,127 9.74 5
Bronx 58,891 17,764 3.32 2
New York (Manhattan) 54,264 13,677 3.97 2
Shelby (Memphis) 51,488 5,700 9.03 5
Alameda (Bay Area) 50,322 12,566 4.00 2
Marion (Indiana) 41,115 9,908 4.15 3
Philadelphia 37,770 7,056 5.35 4
Orange 37,087 12,116 3.06 2
Harris (Houston) 35,535 11,875 2.99 2
San Francisco 34,381 6,809 5.05 3
Baltimore City 32,603 5,670 5.75 4
Mecklenburg (Charlotte) 32,094 8,786 3.65 2
District Of Columbia 31,682 6,787 4.67 3
Maricopa (Phoenix) 31,510 5,948 5.30 4
Contra Costa (Bay Area) 30,255 8,013 3.78 2
Clark (Nevada) 29,906 8,732 3.42 2
Queens 27,355 7,538 3.63 2
Wayne (Detroit) 26,472 5,916 4.47 3
San Diego 26,247 5,693 4.61 3
Total (1,696) 2,741,867 641,841 4.27 2
0
100K
200K
300K
Contact us

ArcGIS Online

The Mapping Platform for Your Organization

Map Your World Your Way

ArcGIS Online gives you everything you need to create interactive web maps and apps that you can share with anyone. With ready-to-use content, apps, and templates you can be productive right away. And no matter what you use—desktops, browsers, smartphones, or tablets—you always have access to your content.

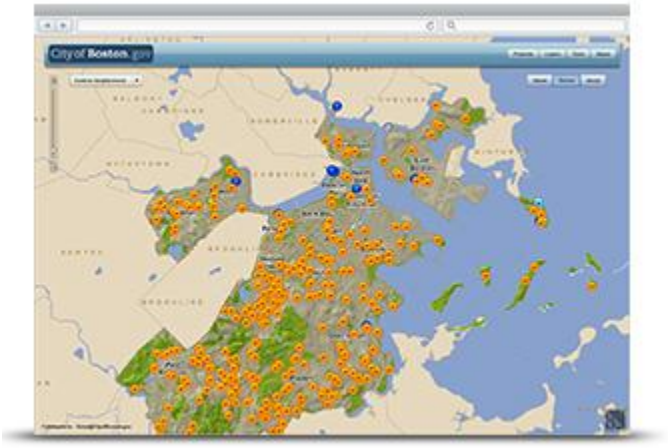
Watch the Video: What is ArcGIS?

Make Maps Your Way

Combine your data with data from Esri and other contributors to create maps for the work you do. Ready-to-use basemaps, tools, templates, and datasets make it easy to design and publish maps online.

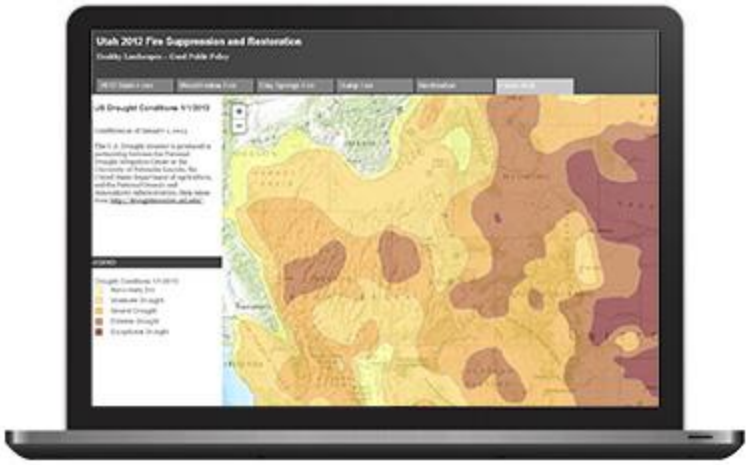
Explore a World of Content

Browse the world’s most extensive online geographic resource and discover maps and data about thousands of topics. Combine content any way you want and see it on a map.



Get Location-Based Insights

Whether you are a scientist, decision-maker, or just curious, you will see location in a new way. By analyzing relationships and patterns you can better understand problems and locate opportunities.



Connect with People Using GIS

Everyone in your organization can share the latest maps, data, and ideas on a single platform. Create project groups, customize work tools, and build your own web apps to connect people through GIS.